Dear Reviewer, we are very grateful for the careful revision of our manuscript and for making our study more complete. We benefited greatly from your feedback and, after careful revision and implementation of your comments and critiques, we are coming back with the manuscript revision and detailed replies to your comments. Please let us know in case of any questions or concerns regarding the new version of the manuscript or the replies below. Our replies are in red and your comments are in black.

This manuscript presents approaches and results very similar to previous studies without relevant increase of information. The concepts and ideas are already well established. Along the manuscript, necessary details and quality controls are missing, especially regarding the numerical reservoir modelling, which prevents from being confident in the quantified results. Moreover, the deterministic THM modelling remains very generic and therefore delivers common knowledge and general results. The scientific significance is missing over the manuscript. Besides, the manuscript is wordy and extrapolates the results. Consequently, I recommend that this manuscript be rejected for publication in Solid Earth and any other journal.

The methodology and the results presented have been discussed in a very similar way and for the same area by the same main author in:

- Kruszewski, M., G. Montegrossi, T. Backers, and E. H. Saenger, 2021, In Situ Stress State of the Ruhr Region (Germany) and Its Implications for Permeability Anisotropy: Rock Mechanics and Rock Engineering, **54**, 6649–6663.
- Kruszewski, M., G. Klee, T. Niederhuber, and O. Heidbach, 2022a, In situ stress database of the greater Ruhr region (Germany) derived from hydrofracturing tests and borehole logs: Earth System Science Data, **14**, 5367–5385.
- Kruszewski, M., G. Montegrossi, M. Balcewicz, G. de Los Angeles Gonzalez de Lucio, O. A. Igbokwe, T. Backers, and E. H. Saenger, 2022b, 3D in situ stress state modelling and fault reactivation risk exemplified in the Ruhr region (Germany): Geomechanics for Energy and the Environment, **32**, 100386.

Despite the probabilistic approach was applied in this manuscript, the added-value is questionable.

Our study investigates in greater detail the probability for fault reactivation utilizing new probabilistic approaches based on known geological uncertainties from the region including the recent comprehensive stress database from the Ruhr region with more than 420 stress magnitude measurements and detailed fault maps created from an abundance of geological data from the coal mining. Such number of geological parameters as well as the new approach to the probabilistic analysis with three scalar values and their thresholds makes our analysis unique. Additionally, to the probabilistic analysis of slip tendency, dilation tendency, and fracture susceptibility, we use the slip vs dilation parameter space to quantify the failure modes of different faults in the region. On top of that, we introduce new scalar value for an improved prospecting of structurally-controlled geothermal resources, called reduced-risk dilation tendency, that allows for quick scanning for structures less likely to slip in shear and more likely to dilate and serve as fluid conduits. Additionally, we use numerical modelling approaches to quantify the evolution of slip and dilation tendency of a fault during long-term geothermal production, quantifying also the change of the failure mode of the fault patch. These components are new, innovative, and have not been yet tackled in the scientific literature as well as were they not part of any of the abovementioned studies mentioned by the reviewer. We consider all the mentioned components of our study to present valuable contribution into the field of reservoir geomechanics and geothermal prospecting in greenfield areas.

2. Regarding the probabilistic approach, why the procedure presented by Seithel *et al.* (2019) (a paper your refer to) is not used although it was developed for the same purpose?

We use (and also update) the methodology and open access *Python* code developed by *Healy and Hicks (2022)* and not the methodology developed in *Seithel et al. (2019)*. The main advantage of the methodology developed in *Healy and Hicks (2022)* is that it is based on a combined Monte Carlo, response surface methodology, and Mohr-Coulomb theory, where the three scalar values (Ts, Td, Sf) for the fault stability analysis are computed for a set of multiple fault segments simultaneously accounting for known geological uncertainties within the reservoir. With their methodology a large amount of simulations can be carried out for a single fault. The results from Python code by *Healy and Hicks (2022)* can be easily transported into maps of probability, like presented in our manuscript. We find, therefore, approach by *Healy and Hicks (2022)* to more applicable for our study.

3. The slip tendency results (Fig. 4) highlight fault patches that have Ts higher than 1. How can it be?

The probabilistic approach considers all model uncertainties which result for extreme cases in faults that are Ts > 1. This would imply that the fault is at failure and should dissipate accumulated elastic stress by failure (either seismic or aseismic slip). However, such a process is not part of the model. The situation of Ts greater than 1 has occurred in a very few cases where normal stress exerted on the fault approached the value of the shear stresses on a fault. It is due to the very few Monte Carlo simulations having large sigma1 values but very low sigma3 values (which are few of the data outliers from the assumed data distribution presented in Figure 2 and then on CDF plots in Figure 3a and b). It should be also mentioned that for the further analysis (maps of fault stability etc.), we take into consideration the probability of Ts exceeding 0.6 and not the few data outliers approaching 1.0. In reality, during e.g., geothermal production Ts will approach or even exceed 1.0 due to the significant lowering of the normal stresses resulting from poro-thermo-elastic effects (as presented in Figure 9a). In short, these are few data outliers resulting from the assumed distributions of principal stresses that have no strong impact on the simulation results.

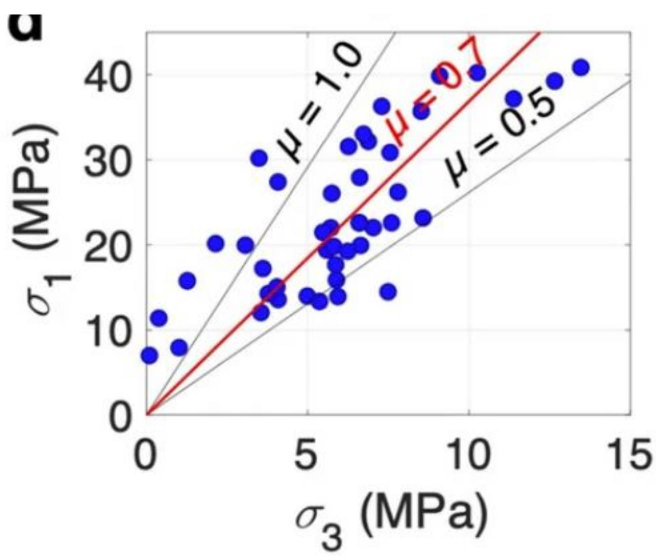
4. With the presented results, many faults should be already critically stressed. How do you explain that no natural seismicity is observed in the area? A chapter discussing the natural seismicity of the area is missing.

In the introduction as well as in the discussion part of the manuscript, we briefly discuss the issue of seismicity in the Ruhr region (L51-53). The seismicity is primarily anthropogenic and connected to either quarry blasting or coal mining (and the recent mine flooding activities). The greater Ruhr region can be considered, therefore, as an aseismic or seismically quiescent region. In the discussion part, we discuss the recent seismic events related to the mine flooding activities in the old coal mines, where relatively small amount of pressures (1 MPa) allowed to create large seismic events of ML 2.6, which we believe to be related to the major fault structures (L193-198). This could indicate that the faults (NW-SE striking) in the region are either critically stressed or close to being critically stressed. The lack of seismicity in the region can be, however, explained by e.g., a release of seismic energy with aseismic fault movement i.e., fault creep. More studies are, however, need to prove this theory.

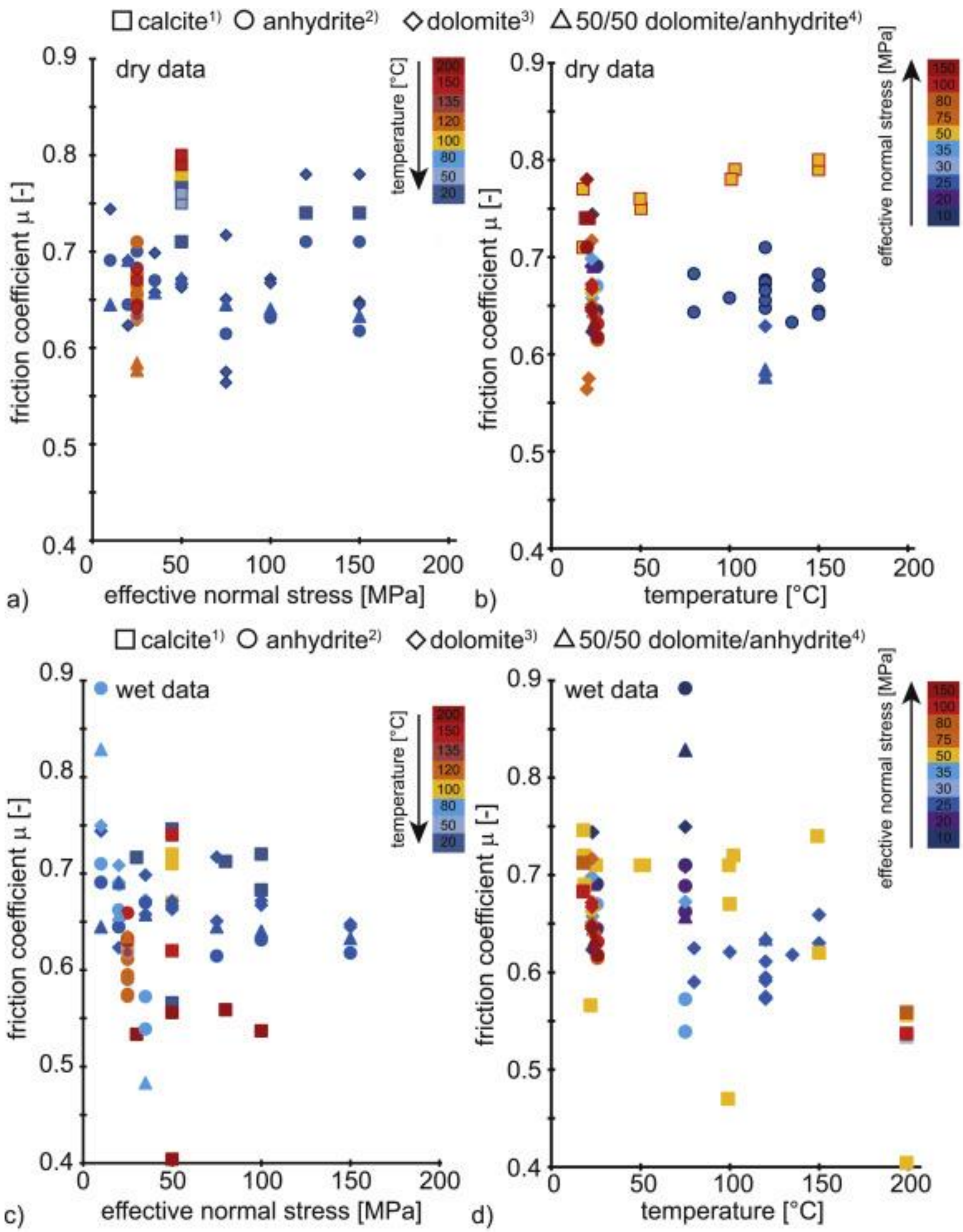
5. For the range of pressure found in the Ruhr area at that depth (<2 kbar), Byerlee (1978) observed friction coefficients of 0.85. However, the limit assumed in the manuscript is 0.6 (L132, L187, Table 2) and not 0.85, using the same reference, why? L210, however, refereeing to Byerlee (1978) again, the 0.85 friction is written to be a possible value!

The nearly 50-year-old *Byerlee (1978)* paper has included static friction coefficient of many different rock types carried out using different types of laboratory testing techniques. The function of static friction coefficient introduced by Byerlee is rather a rough approximation of rock friction independent of rock type or temperature and pressure conditions. We decided to use the static coefficient of 0.6, as suggested by e.g., *Zoback (2009)* and many other scholars in reservoir geomechanics. The assumption of static friction coefficient of 0.6 is, however, not at all random, but it was based on indications of static frictional properties of rocks i) in the Ruhr region and ii) carbonate rocks. We show below examples from the literature of static friction coefficients computed based on either field or laboratory data. Based on the studies below, it can be seen that the static friction coefficient will rarely exceed 0.85 and will be in range between 0.6 and 0.7. This means that the frictional properties of rock will depend significantly on the type of rock as well as on the temperature and pressure conditions expected in the reservoir. We, therefore, still believe that the assumption of static friction coefficient of 0.6 for the Devonian carbonates, made in our study, is appropriate. Laboratory tests on samples from the region, however, are needed to prove this value.

In L162-163 (in the new version of the manuscript) as well as in Figure 3 we merely show the allowable range of static friction coefficient as indicated by *Byerlee (1978)* and not the friction value assumed in this study. The assumed value of 0.6 is indicated in the text (L162) and in Table 1. We add to the Table 1 the references included below. Below we include static friction coefficient estimated from hydrofracturing tests in the Ruhr region (performed at depths of coal mines in the region until approximately 1.4 km depth) from *Kruszewski et al. (2022)*:



Below a figure from *Pluymakers et al. (2016)* showing static friction coefficient for different carbonate rocks.



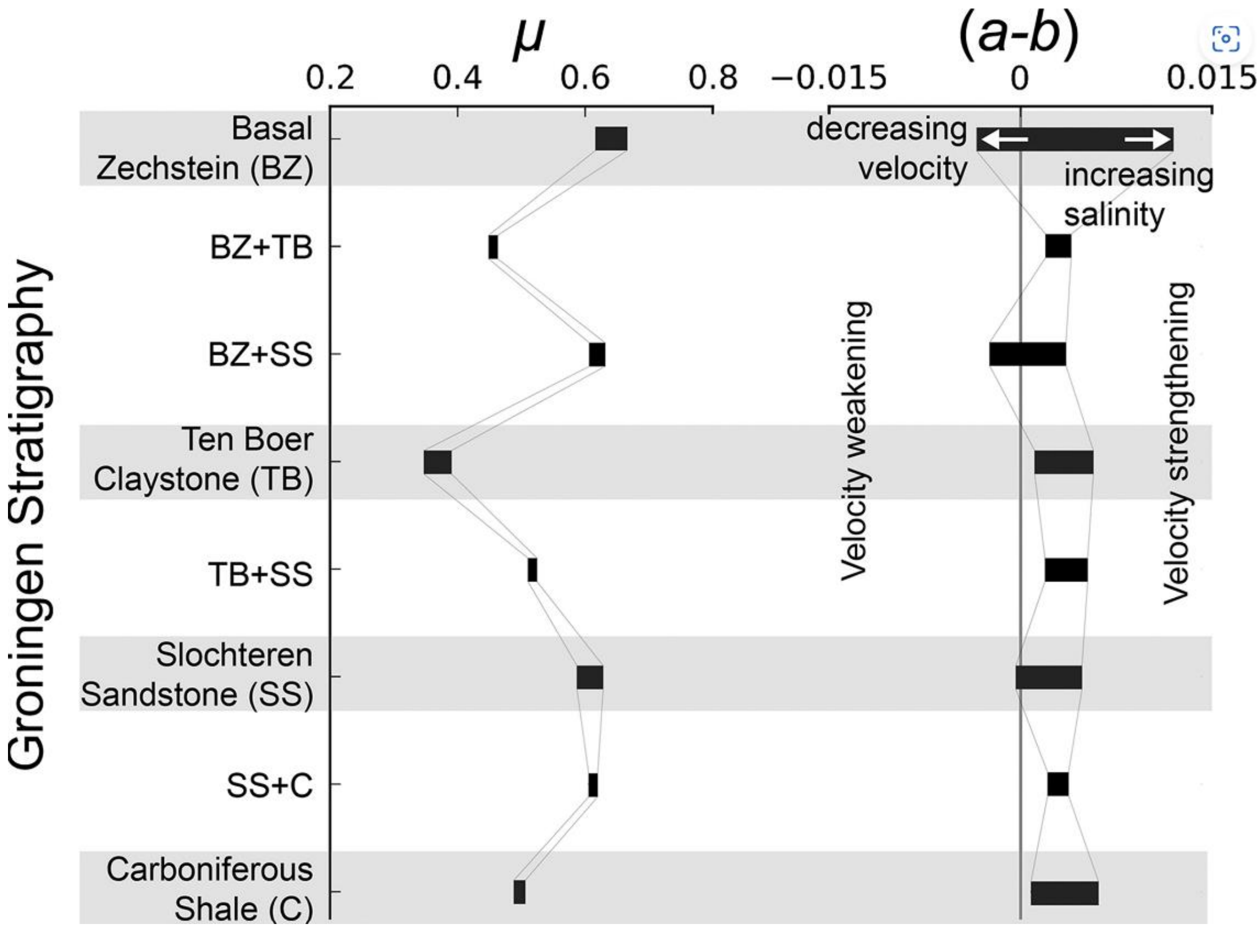
1) Morrow et al, 2000; Shimamoto and Logan, 1981; Verberne et al, 2010, 2013, 2015

2) Scuderi et al, 2013; Shimamoto and Logan, 1981; Pluymakers et al, 2014; Pluymakers and Niemeijer, 2015; Pluymakers et al, current study.

3) Scuderi et al, 2013; Shimamoto and Logan, 1981; Pluymakers et al, current study; Weeks and Tullis, 1985; Samuelson et al, pers.comm.

4) Scuderi et al, 2013, Pluymakers et al, current study.

Below a figure from *Hunfeld et al. (2017)* showing static coefficient of friction for rocks in the Groningen field in the Netherlands.



6. For Sf, in Eq. 3, Co is accounted for, but it does not appear for the slip tendency although both parameters (Ts and Sf) have the same theoretical background (Mohr-Coulomb failure criterion). Why is it so?

We agree with the comment. We remove cohesion from the analysis and from the manuscript focusing on an idealized case of a cohesionless fault(s). The main reason from that is that the available data on cohesion from the region is, as of now, nonexistent. As a result, we amend calculations of Sf and include these changes in the manuscript. Due to this, we provide new maps of Sf in the revised version of this manuscript.

7. In section 5.2, L378-379: "*Scalar values used for fault stability evaluation based on the contribution of fluid pressure only, such as Sf, will not provide a full picture of the fault stability in situ*". This is also true for the slip tendency, so mention it as well.

We agree with the comment and make appropriate changes to the manuscript by removing the sentence.

8. The last sentence of the conclusion is not surprising and does not need any result of the numerical simulation that was described in the manuscript. Below is a (non-exhaustive) list of papers that are already

discussing the importance of thermally induced stress changes on a long term basis in geothermal contexts:

- De Simone, S., V. Vilarrasa, J. Carrera, A. Alcolea, and P. Meier, 2013, Thermal coupling may control mechanical stability of geothermal reservoirs during cold water injection: Physics and Chemistry of the Earth, Parts A/B/C, **64**, 117–126.
- Egert, R., Gaucher, E., Savvatis, A., Goblirsch, P., Kohl, T., 2022. Numerical determination of long-term alterations of THM characteristics of a Malm geothermal reservoir during continuous exploitation. Presented at the European Geothermal Congress 2022, Berlin, Germany.
- Jeanne, P., J. Rutqvist, and P. F. Dobson, 2017, Influence of injection-induced cooling on deviatoric stress and shear reactivation of preexisting fractures in Enhanced Geothermal Systems: Geothermics, **70**, 367–375.
- Jeanne, P., J. Rutqvist, P. F. Dobson, J. Garcia, M. Walters, C. Hartline, and A. Borgia, 2015, Geomechanical simulation of the stress tensor rotation caused by injection of cold water in a deep geothermal reservoir: Journal of Geophysical Research: Solid Earth, **120**, 8422–8438.
- Kivi, I. R., E. Pujades, J. Rutqvist, and V. Vilarrasa, 2022, Cooling-induced reactivation of distant faults during long-term geothermal energy production in hot sedimentary aquifers: Scientific Reports, **12**, 2065.
- Koh, J., H. Roshan, and S. S. Rahman, 2011, A numerical study on the long term thermo-poroelastic effects of cold water injection into naturally fractured geothermal reservoirs: Computers and Geotechnics, **38**, 669–682.
- Wassing, B. B. T., T. Candela, S. Osinga, E. Peters, L. Buijze, P. A. Fokker, and J. D. Van Wees, 2021, Time-dependent Seismic Footprint of Thermal Loading for Geothermal Activities in Fractured Carbonate Reservoirs: Frontiers in Earth Science, **9**.

Many references regarding THM modelling in similar contexts should be given but they are missing. They could have been used as inspiration source.

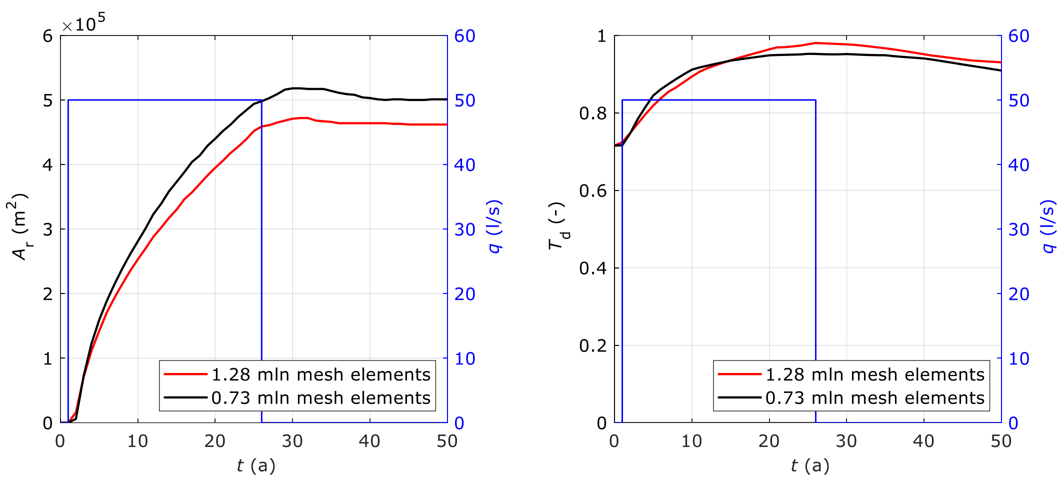
We have added few of the mentioned references to the manuscript (L67-68). We have changed the rationale of the study, where now we do not discuss the thermal effects on fault reactivation but rather focus on the temporal evolution of slip and dilation tendencies of the fault during long-term geothermal production to show the evolution of the reactivation potential in time (and space) as well as the change of the fault failure conditions on the Ts vs. Td parameter space. Thermal effects, as pin-pointed by the reviewer, is a widely known phenomenon already discussed in the literature and, therefore, we decide to not discuss it in the new version of the manuscript.

9. When presenting THM numerical modelling, it is necessary to develop much more what is actually done to give confidence in the results. So far, it is not the case, and a lot of information is missing, e.g. what are the physical processes activated (equations)? The above-mentioned papers could help to do so.

As already mentioned, we change the rationale of our study, where numerical modelling is not anymore main part of the methodology and not main part of the study. Numerical model is now included just in the discussion part of the manuscript (Section 5.4) and is used only to support the arguments stated in the main parts of the study. We decide, therefore, not to discuss in extensive detail all the physical processes and equations used in the numerical model as we deem it redundant. We describe what processes control the developed numerical model, what are the input parameters and what are the boundary conditions used, both in the text (Section 5.4; L285-307) and in improved Figure 8. In the *Data Availability* section of the manuscript we publish (open access) the developed numerical models (see *Kruszewski and Verdecchia (2023)* in the references), including all necessary information about physical processes, equations, input parameters, boundary conditions, discretization etc., with the manuscript to allow reproducibility of the results presented in our study. The updated numerical models will be published with the revised version of this manuscript. These models can be easily checked by readers or reviewers for reproducibility. For more detailed explanations of the physical processed/equations used in the numerical model we refer the reviewer to *COMSOL (2021)* as well as *Taillefer et al. (2018)*, both are referenced in the manuscript.

10. Was a mesh sensitivity study carried out? This is questionable when looking at the discontinuous curves of Fig. 7a and b.

Yes, we have carried out convergence tests on models three different mesh sizes. Below a snapshot comparing results of the cumulative reactivated fault area (Ar; on the left) as well as the maximum dilation tendency on the fault plane (on the right) computed with the model discretized into 0.73 $\cdot 10^6$ elements and one discretized in 1.28 $\cdot 10^6$elements. We skip the model with 0.3 $\cdot 10^6$ elements, where model results were deemed to be dependent on mesh size. The results we deemed to be satisfactory and we use the model discretized in 1.28 $\cdot 10^6$ elements for our discussion in the manuscript.



11. In the THM results, it would be most important to see space and time distribution of, at least, the pore-pressure field and the temperature field before jumping directly to the shortest distance between wells and fault.

We make changes to our approach and we use now numerical modelling only in the discussion part of the study to show the evolution of both slip and dilation tendencies as well as to compute the cumulative reactivated fault area and fault area with dilation tendencies larger than 0.8. We decided not to discuss the thermal and pore pressure effects at all in the manuscript. This, as pin-pointed by the reviewer, is widely known phenomenon already discussed in the literature. We decide, therefore, to not include the time and space evolution of pressure and temperature fields as advised by the reviewer.

12. Section 3.2, L165-168: "*Effects such as fault permeability enhancement due to the dilation, change of rock properties due to Pp or temperature, T, the influence of fluid chemistry on rock mass and fault properties, mechanisms of earthquake interactions, and the Kaiser effect are not considered in the simulation*" This looks like COMSOL could account for all of these. I am not aware that COMSOL can simulate earthquakes.

We do not account for the listed effects in our study and it is mainly due to the lack of published data from the region on fault permeabilities, change of rock properties with T,P conditions, fluid chemistry, Kaiser effect, earthquake interaction and so on. Performing simulation with all listed effects will lead to a significantly overconstrained model with much larger uncertainties, which was not our aim in this study. With our analysis, using simple and idealized numerical models, based on limited input and geological data from the geothermal reservoir, we investigate the possible fault reactivation and the evolution of slip and dilation tendency in time and space. Although, we would like to add all of the listed effects to our work in the future, when more data will become available from the region, as of now, we find that the simulation results presented in our study still give a good picture of what could occur in the subsurface with the amount of data that is, as of now, available.

13. First sentence of abstract: This is wrong as underlined e.g. by the deep geothermal exploitation in the Paris basin for many decades.

We agree with the comment. We make it more precise now and include a part where we say that our analysis tackles only structurally controlled geothermal systems (where the matrix permeability in insufficient for geothermal fluid production; L25-30). This has been made also now clearer with the new title of our manuscript.

14. Nothing in the manuscript supports the simulation of seismicity or aseismic slip or seismic hazard. Consequently, these aspects should be mentioned with care.

We agree with the comment and make appropriate changes to the manuscript.

15. Second sentence of abstract: what is the Earth's "plumbing" system? I have never read such wording in a geothermal context. Do you mean "circulation"?

We agree with the comment and change the phrasing in the manuscript. By *plumbing system*, we meant the *hydraulic system*, but we agree that it was rather colloquial phrasing (L1).

16. L18: […] a complex "web" of faults […]? I have never read such wording in a geothermal context. Do you mean "network"? It is found L401 as well.

We agree with the comment and change the phrasing in the manuscript. We meant fault network.

17. Avoid using "the distance to failure" (e.g. L31), you mean in meters (?), prefer the "reactivation potential".

We agree with the comment and change the phrasing in the manuscript.

18. The Appendix does not correspond at all to what is announced in the main part of the manuscript.

We agree with the comment. We integrate the appendix into the main text and amend the text accordingly (L115-136).